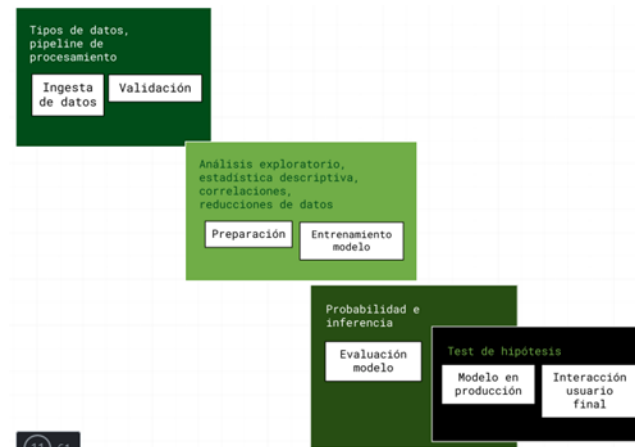


code002

June 28, 2024

# 1 Flujo de trabajo en Ciencia de Datos



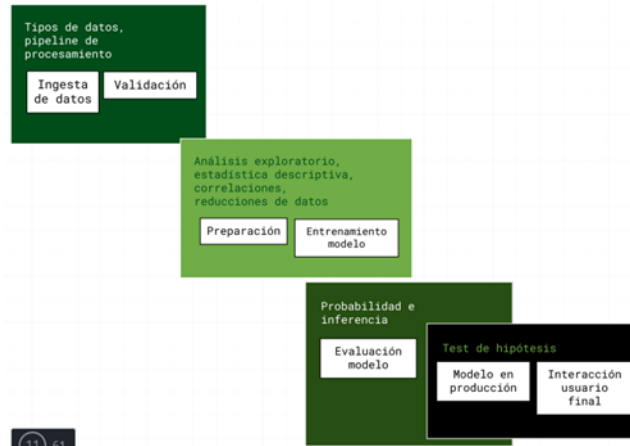
## 1.1 Primer fase

- Data ingestion (Ingesta de datos)
- Data Visualization (Visualización de datos)
- Data preparation (Preparación de los datos) ## Segunda fase
- Model training (Entrenamiento del modelo)
- Model evaluation (Evaluación del modelo)
- Model validation (Validación del modelo) ## Tercera fase
- Model serving
- End user interface (Entrega a usuario)

Existe una secuencia recomendable, la primera fase consiste en recabar la información y acondicionarla para que obtengas datos útiles para que sea la entrada perfecta a lo que será el modelo; luego aquí vienen las subfases de entrenamiento de modelo, evaluación y validación del modelo. Cuando esa parte está cubierta, el modelo ya puede pasar a producción y luego termina con el usuario final.

Hay diferentes roles para cada tarea dentro de este proceso así que cada perfil está más afin o enfocado en alguna tarea específica.

Pero todos necesitan colaborar y claridad sobre que en algún punto cada perfil va a necesitar del uso de Estadística dentro de todo el flujo de trabajo.



- **Ingesta de datos y validación:** Se refiere al preprocesamiento e identificar en un dataset; *tipos de datos y el pipeline de procesamiento (flujo de procesamiento)*. Ese flujo de procesamiento va a definir cuales son las transformaciones que yo debo hacer sobre los datos, para que así con esas transformaciones los datos queden limpios y adaptados a lo que necesita el modelo, que yo como científico de datos tengo claro que el modelo va a ejecutar para resolver un problema puntual de negocio.
- **Preparación y entrenamiento del modelo:** Esta es la fase de terminar de procesar los datos y empezar a contruir el modelo. Aquí vamos a encontrar elementos interesantes como; Analisis exploratorio de los datos y uso de estadística descriptiva en el analisis, entender si existen correlaciones en los datos y con base en eso, también hacer posibles reducciones de datos.
- **Evaluar, mandar el modelo a producción y entrega a usuario:** Aquí hay elementos de la estadística como es probabilidad e inferencia, aquí vamos a usar mucho algo que se usa para probar modelos “el test de hipótesis”. Todos estos elementos hacen parte mas de la rama de estadística inferencial.

Los 2 bloques iniciales tienen mayor tendencia de uso de la estadística descriptiva y el último se inclina más por la inferencial.

## 1.2 Extras:

[Acelerar el flujo de trabajo BBVA](#)